Satellite-based PM concentrations and their application to COPD in Cleveland, OH

Naresh Kumar[1], **Dong Liang**[2], **Alejandro Comellas**[3], **Allen D. Chu**[4], and **Thad Abrams**[5]

[1]Department of Public Health Sciences, University of Miami, Miami, Florida, USA

[2]Department of Epidemiology, University of Iowa, Iowa City, Iowa, USA

[3]Department of Pulmonary Medicine, University of Iowa, Iowa City, Iowa, USA

[4]Goddard Space Flight Center, NASA, Greenbelt, Maryland, USA

[5]Iowa City VA Medical Center, Iowa City, Iowa, USA

Abstract

A hybrid approach is proposed to estimate exposure to fine particulate matter ($PM_{2.5}$) at a given location and time. This approach builds on satellite-based aerosol optical depth (AOD), air pollution data from sparsely distributed Environmental Protection Agency (EPA) sites and local time–space Kriging, an optimal interpolation technique. Given the daily global coverage of AOD data, we can develop daily estimate of air quality at any given location and time. This can assure unprecedented spatial coverage, needed for air quality surveillance and management and epidemiological studies. In this paper, we developed an empirical relationship between the 2 km AOD and $PM_{2.5}$ data from EPA sites. Extrapolating this relationship to the study domain resulted in 2.3 million predictions of $PM_{2.5}$ between 2000 and 2009 in Cleveland Metropolitan Statistical Area (MSA). We have developed local time–space Kriging to compute exposure at a given location and time using the predicted $PM_{2.5}$. Daily estimates of $PM_{2.5}$ were developed for Cleveland MSA between 2000 and 2009 at 2.5 km spatial resolution; 1.7 million (~79.8%) of 2.13 million predictions required for multiyear and geographic domain were robust. In the epidemiological application of the hybrid approach, admissions for an acute exacerbation of chronic obstructive pulmonary disease (AECOPD) was examined with respect to time–space lagged $PM_{2.5}$ exposure. Our analysis suggests that the risk of AECOPD increases 2.3% with a unit increase in $PM_{2.5}$ exposure within 9 days and 0.05° (~5 km) distance lags. In the aggregated analysis, the exposed groups (who experienced exposure to $PM_{2.5}$ >15.4 $\mu g/m^3$) were 54% more likely to be admitted for AECOPD than the reference group. The hybrid approach offers greater spatiotemporal coverage and reliable characterization of ambient concentration than conventional *in situ* monitoring-based approaches. Thus, this approach can potentially reduce exposure misclassification errors in the conventional air pollution epidemiology studies.

Correspondence to: Dr. Naresh Kumar, Department of Public Health Sciences, University of Miami, Miami, FL 33136, USA. Tel.: +1 305-243-4854. Fax: +1 305-243-5577. nkumar@med.miami.edu.

INTRODUCTION

Health effects of air pollution have been subject to intensive research scrutiny for the past several decades.[1] Although we have made great strides in understanding the etiology of the health effects (such as inflammation, oxidative stress and DNA damage) of air pollution,[2–7] it has been difficult to quantify the precise burden of disease and disability associated with ambient air pollution, especially for population-based epidemiological studies, for several reasons, including exposure uncertainty and study designs. Exposure uncertainty arises because of disconnect between the spatiotemporal resolution of air pollution and health data, and poor spatiotemporal coverage of air pollution data.[8] Given the mismatch in the spatiotemporal resolution of air pollution and health data, most studies rely on retrospective research design, and aggregate ambient air pollution (at sparsely distributed sites) to assess exposure. Given that air pollution can vary within short distance and time intervals, city- or county-wide aggregated estimates of air pollution cannot truly represent population exposure.[9–11]

Given the limited spatiotemporal coverage of air pollution monitoring, data from centrally located sites are inadequate to quantify precise exposure.[12] Recent advances in satellite remote sensing,[13,14] chemical transport models[15–18] and expansion of *in situ* monitoring network offer an opportunity to develop exposure at unprecedented spatiotemporal resolutions. Although the Environmental Protection Agency (EPA) air pollution monitoring network has been expanded, its spatial coverage is still sparse. Thus, this constrains our ability to adequately estimate population exposure for epidemiological studies. This paper offers a novel hybrid approach to develop time–space resolved estimates of exposure. The proposed approach builds on the strengths of high-resolution satellite remote sensing data, EPA data from *in situ* monitoring of air pollution and local time–space Kriging (LTSK), an optimal interpolation program we have developed.[19] The proposed approach offers several advantages over alternate measures of quantifying exposure. First, Terra and Aqua satellites have daily global coverage and data from sensors aboard these satellites, especially MODIS (Moderate Resolution Imaging Spectroradiometer), can be used to retrieve AOD at greater spatial resolution (i.e. as fine as 1 and 2 km) for two times a day (~1030 and ~1330 hours);[14,20] average of these two values: close to peak hours (~1030 hours local time) and off-peak hours (~1330 hours local time) can provide daily estimates of AOD. In addition, satellite data are in public domain and easily accessible to researchers and regulatory authorities. Therefore, the proposed approach has the potential to enhance the spatiotemporal coverage of air pollution estimates. Second, the hybrid approach addresses the problem of mismatch in the spatiotemporal scales of health and air pollution data. Although the empirical model between satellite-based AOD and EPA data sets are used to develop air quality estimates, LTSK resolves the problem of mismatch in the spatiotemporal scales of air pollution and health data sets.[19] LTSK can be used to quantify exposure at the (desired) spatiotemporal scales of health data sets. Third, the hybrid approach utilizes *in situ* data (from EPA sites) for developing empirical model. Thus, the model performance and model predictions can be validated. Given the spatiotemporal heterogeneity in emission sources and meteorological conditions, the hybrid approach can be used to develop region-specific models to predict air pollution within the region-specific domains.

The hybrid approach was used: (a) to develop daily estimates of airborne particulate 2.5 μm in aerodynamic diameter (PM$_{2.5}$) at 2 km spatial resolution from 2000 to 2009 for the Cleveland Metropolitan Statistical Area (MSA), and (b) to compute time–space lagged exposure to PM$_{2.5}$ for veterans admitted for an acute exacerbation of chronic obstructive pulmonary disease (AECOPD) in Cleveland MSA. The remainder of this paper describes the study area, materials and methods used for developing high-resolution estimates of air pollution and quantifying time–space lagged exposure with the aid of LTSK and risks of AECOPD with respect to time–space lagged exposure to PM$_{2.5}$. The last two sections present results and a discussion on the implications of the hybrid approach for air pollution epidemiological studies and air quality surveillance and management.

MATERIALS AND METHODS

Study Area

The study focuses on Cleveland MSA, which has diverse urbanization and several EPA sites that monitor airborne particulates <2.5 pm in aerodynamic diameter (PM$_{2.5}$) (Figure 1). The *in situ* monitored data from these sites are important for developing and validating statistical model to estimate PM$_{2.5}$. The study area also includes both contrasting urban downtown (site no. 60) and suburban areas (site no. 17) and diverse emission sources. Cleveland has consistently failed to meet EPA thresholds of coarse particles. Moreover, several studies document adverse health risks (such as hospital admission and mortality because of cardiopulmonary disease) of ambient particulate matter.[21–23] The annual average temperature in the study area is 19 °C and ranges from 3 °C to 4 °C in December–January to 27 °C in July–August. A vast majority of time, the wind direction between 0800 and 1500 hours is southwest.

Data Sources

This research relies on data from four different sources, namely MODIS Level 1 and Level 2 data from NASA,[24] meteorological data from the National Climatic Data Center,[25] air pollution data from EPA[26] and COPD data from Veteran Affairs (VA) in-patient and outpatient administrative databases. These data are described below.

MODIS data—Terra and Aqua satellites (that have MODIS onboard) were launched on 18 December 1999 and 4 May 2002, respectively, and MODIS data have been available since 24 February 2000 and 24 June 2002, respectively. MODIS records spectral radiances in 36 bands, which can be grouped by three different spatial resolutions: 0.25, 0.5 and 1.0 km. Although NASA has developed algorithm to extract AOD, the finest spatial resolution of the AOD data is 10 km. Thus, these products were of little use for this research. To extract AOD at finer resolution, we acquired the following MODIS data sets through December 2009 from both satellites: Level 1b Calibrated radiances—1.0 km; Level 1b Calibrated radiances—0.5 km; Level 1b Calibrated radiances—0.25 km; Geolocation—1.0 km; Level 2 Join Atmospheric Products of Profiles; Total Column Ozone; Water Vapor and Stability Indices; and Level 2 Cloud Mask and Spectral Test Results. We modified PGE04 (Version 5.1.0)[27] to extract 2 km AOD using these data. Details on the 2 km AOD extraction are described in the Methodology section below (see Kumar et al.[28] for details).

Meteorological data—Global surface hourly data on meteorological conditions, including relative humidity, surface temperature, wind direction, wind speed, dew point and atmospheric pressure, were acquired from National Climatic Data Center. These data were important for developing an AOD–PM$_{2.5}$ empirical model, because meteorological conditions can influence AOD greatly. For example, the value of AOD increases with the increase in relative humidity, not only because it increases the concentration of water vapors

but also because it inflates particle size.[29] Other factors, such as wind speed and atmospheric pressure, can influence aerosols mixing within the boundary layer height.[30,31] This, in turn, also influences uncertainty in AOD retrieval.

EPA data—$PM_{2.5}$ data from 2000 to 2009 were acquired from EPA from all monitoring stations in the Cleveland MSA[26] (Figure 1). These data were needed to develop an empirical PM–AOD model and evaluate predicted $PM_{2.5}$. Although there were many monitoring stations, data from five $PM_{2.5}$ were used, because data from only these stations had adequate number of data points within 0.025° (~2.5 km) distance and 60 min time intervals of AOD data.

Hospitalization data—Data on hospitalizations include all in-patient admissions for an AECOPD among veterans to the US Department of VA facilities within the Cleveland MSA during 2007. These data were acquired from VA administrative databases and included: the Patient Treatment File; Outpatient Care File; Vital Status File; and the Planning Systems Support Group Enrollment File. All databases have unique identifiers comprised of formula-based encryptions of individuals' Social Security Numbers. The identifier is consistent for a given patient across all data sets and fiscal years. In the fiscal year 2007, a total of 369 veterans were admitted with a primary diagnosis of an AECOPD and their geocoded addresses (with XY coordinates, that is, longitude and latitude) were available.

Methodology

We extracted AOD at 2 km spatial resolution, which is more robust than the coarse resolution AOD. The methodology for 2 km AOD retrieval and its comparison with the coarse resolutions (5 and 10 km) of AOD is described elsewhere.[28] AOD, which represents ambient solid and liquid particles suspended in the air, can be influenced by several factors, including AOD retrieval from satellite data, such as cloud contamination, surface glint, types of aerosols and boundary layer height.[9,14,32] Nonetheless, AOD retrieval at a greater spatial resolution, such as 2 km used in this research, can help overcome some of these problems, because smaller the geographic area lesser the cloud contamination and reflectance uncertainty because of less heterogeneity in land-use and land-cover types. Kumar et al.[33] show that the 2 km AOD (retrieved using MODIS data) shows better association (coefficient of correlation ~0.93) with *in situ* measurement of AOD at Aeronet sites[33] than the 10 km AOD (correlation of coefficient ~0.76).

It is important to note that AOD retrievals are not possible every day because of cloud cover, cloud contamination and distortion in surface radiation due to snow cover. Between 28 February 2000 and 31 December 2009, AOD retrieval was possible for 2176 (60%) days. We chose not fill these gaps using the imputation technique for two reasons. First, imputing values to fill these gaps and then using these imputed and actual AOD values combined to estimate exposure (at a given location and time) will result in greater bias in prediction because of error propagation at two levels. Second, if we have robust estimates of $PM_{2.5}$ for days, we have AOD retrievals, and LTSK can be used to estimate $PM_{2.5}$ at any given location and day (using these data with the gaps).

Using the *in situ* data monitored at EPA sites, an empirical relationship between $PM_{2.5}$ and 2 km AOD was developed, which controlled for meteorological conditions, seasonality and temporal structure. Because AOD can greatly be influenced by meteorological conditions, we instrument AOD on meteorological conditions using the *xtivregress* function in STATA.[34] This function controls for site- and day-specific error structure in the data. The model estimates were extrapolated for the entire study domain, which included 2.3 million predictions of daily $PM_{2.5}$ between 2000 and 2009 in Cleveland MSA (geographic extent

82.5 °W to 80.75 °W and 40.5 °N to 41.92 °N). There were only five sites in the study area and generalizability of the site-specific method has to be restricted within small area around the site. Thus, we use global model. As evident from Table 2b, the difference between observed and predicted mean in the cross-validation (for all sites) is <1 μg/m^3. The global model was run for all sites, but the statistics presented in Table 2b include site-specific summaries. The performance of the global model should be assessed based on the summary of estimates from all sites. The difference between observed and predicted mean is not significantly different as compared with the site-specific model.

Spatiotemporal scales of predicted PM$_{2.5}$ data are not the same as the spatiotemporal scales of health data. We have developed LTSK to address this mismatch in the spatiotemporal scales of two different data sets. If sufficient sample data are available, LTSK can estimate the value at a given location and time within the spatiotemporal domain of the data set. LTSK offers several advantages over the classical time–space Kriging. First, environmental data sets exhibit non-stationary structure, for example, sources of air pollution vary within a region. Therefore, covariance structure within a region is unlikely to be the same. Likewise, interactions between emission sources and meteorological conditions generate processes that are non-stationary both in the mean and covariance structures. In LTSK, we estimate model parameters for each (given) query point for which prediction is needed. The model parameters are estimated using the data points within the selected spatiotemporal intervals (or window or voxel) around a given query point (for which exposure needs to be estimated). The proposed LTSK is a non-separable product–sum model adopted from De Iaco et al.[35]

Let $\{z(s,t)\}$ denote the observed Gaussian process defined over $D \times \{1,2,\ldots,T\}$, where $D \subset \Re^2$ denotes the spatial (geographic) domain and t indexes discrete time stamps (such as day in our case). Let (s_j,t_j), $j = 1,\ldots,m$ denote the location and time where predictions are needed. We term (s_j,t_j) as a query point. The implementation of local Kriging requires a location-specific neighborhood around a query point (s_j,t_j), denoted by N_j. We assume that within the local neighborhood, data are second-order stationary and observe non-separable spatiotemporal covariance function, specified using the product–sum approach.[35] Let H and U denote the distance and time intervals for defining neighbors. These spatiotemporal intervals should be chosen based on some prior knowledge. For example, for AOD this knowledge can be based on the life cycle of the aerosol and spatial resolution of the satellite data. We define a cylinder N^*_j around each query point as

$$N^*_j \equiv \{(s_k,t_k)|\ \|s_k-s_j\| <H\ and\ |t_k-t_j| <U\} \quad (1)$$

The strength of the underlying spatiotemporal correlations is allowed to vary both spatially and temporally around query points. We estimate a location-specific variogram $\hat{\gamma_j}(h,u)$ with the procedure described below. Given the estimated spatiotemporal range parameters, denoted by $\hat{h}_j$ and $\hat{u}_j$, respectively, we define the local neighborhood as a subset of N_j^*:

$$N_j=\{(s_k,t_k)|\ \|s_k-s_j\| <\hat{h}_j\ and\ |t_k-t_j| <\hat{u}_j\} \quad (2)$$

We thus identify a local spatiotemporal domain, where the process at observed (or sample) locations is highly correlated with the process realized at query point (j), conditioning on the estimated variogram.

We estimate the empirical variogram $\hat{\gamma_j}(h,u)$ using data points within H geographic distance and U days around the query point N_j^*. We used the method of moment estimates to choose

spatial and temporal variograms among exponential, spherical, Gaussian and Matern models.[36] Range parameters can be estimated automatically using the initial lag in the 80 percentile of the spatial- or temporal-only variograms.[37] Nugget and sill parameters are estimated using weighted least square, adjusting for positive correlation within local neighborhood.[38] The interaction parameter of the product–sum model is adjusted to achieve a negative definite estimate of the variogram around each query point.

In implementing LTSK, we face two challenges: sparse data and computation time. First, narrow spatiotemporal lags result in very few data points to accurately estimate time–space covariance function.[39] Consequently, it results in discontinuity in the prediction across time and space. For example, in the satellite-based air pollution data, there are systematic gaps (due to cloud cover and/or data contamination) and sufficient data points may not be available if small time–space lags are used to define local neighborhood. The second challenge is the computation of empirical variogram and the Kriging operation, which involves inversion of a dense $n_j \times n_j$ matrix, where n_j denotes the number of neighbors around a query point j. If n_j is large, repetitive inversion requires $\mathrm{o}(mn_j^3)$ computation and the implementation becomes computationally expensive (sometimes prohibitive) for a large number of m

We utilize dimension reduction and subsampling techniques to address these challenges.[40–42] When the number of neighbors around a query point (n_j) exceeds a fixed limit (L_n), we sample L_n data points in a spatiotemporally balanced way. A balance sampling achieves a good predictive performance of the underlying process.[43] This requires setting up non-overlapping voxels (or cubes) around the query point j and drawing a weighted sample within each voxel, with probability proportional to the number of data points within each voxel. If there are less than three distinct non-empty cubes across space or time, variogram estimation and Kriging operation are not performed and warning messages ("Less Neighbors by Space" or "Less Neighbors by Time") are attached with the output; more details on the theoretical model and implementations on LTSK are available in the Supplementary Online Material.

LTSK for exposure estimation—We applied LTSK to two applications: (a) prediction of daily $PM_{2.5}$ at 2.5 km (i.e. the centroid of each 2.5 km pixel) spatial resolution for the Cleveland MSA from 2000 to 2009, and (b) time–space lagged ambient $PM_{2.5}$ exposure at the place of residence for all veterans in the study area admitted for AECOPD during 2007. The average exposure was computed within 3, 6, 9, 12 and 15 days of exacerbation and within 0.025°, 0.05°, 0.075° and 0.1° (or ~2.5, ~5, ~7.5 and ~10 km) distance of residential location of each subject's home address.

Statistical analysis—Home addresses of veterans who were admitted for AECOPD in Cleveland MSA in 2007 were geocoded. Given our AECOPDs are secondary retrospective data and we did not have access to data on potential confounders, we chose the case crossover design. In this design, all subjects are cases and they serve as their own controls. Thus, their exposure is measured at two points in time: first when they have (disease) symptoms (and serving as cases) and second when they are (disease) symptom free (serving as their own controls).[44] The main assumption underlying the case crossover design is that everything concerning each case is the same but the exposure. Thus, it allows us to differentiate the effects of change in exposure on the occurrence of disease. Although the date of occurrence of disease or symptom is recorded and available in the data set, the time when cases serve as their own controls during the disease/symptom-free period needs to be determined with caution. Even though the control period can be chosen in many ways, we used 30 days before the exacerbation, because about 4 weeks before exacerbation these

patients should not have conditions that lead to exacerbation except air pollution. In addition, the chosen time period is not distant. We assume that socioeconomic and demographic conditions were expected to be the same at both points in time. Let $y_{i0} = 0$ indicate no admission (because of AECOPD during the control period) and $y_{i1} = 1$ indicate hospital admission because of AECOPD. Let z_{ij} denote time–space lagged exposure to $PM_{2.5}$ exposure for the selected spatiotemporal lags as outlined above (within 15 days at an interval of 3 days and within 10 km at an interval of 2.5 km). We assume a conditional logistic regression model $y_{ij} \sim Bernoulli(p_{ij})$ as

$$logit(p_{ij}) = \alpha_i + \beta z_{ij} + \varepsilon_{ij} \text{ for } i = 1, \ldots, n \text{ and } j = 0 \text{ or } 1 \quad (3)$$

Two different sets of analyses were performed. In the first set, z_{ij} served as a continuous variable (i.e. exposure to $PM_{2.5}$ measured in $\mu g/m^3$) and relative risks were estimated with respect to a unit change in $PM_{2.5}$ exposure within the selected spatiotemporal lags. In the second set, relative risks of AECOD were estimated across exposed and reference groups. We used 75th percentile of the computed $PM_{2.5}$ exposure to define exposed and reference groups. In the exposed group, $PM_{2.5}$ exposure (within the selected spatiotemporal lags) was >15.4 $\mu g/m^3$ (which is 75th percentile and close to the annual average standard of $PM_{2.5}$ (15 $\mu g/m^3$) as adopted by EPA). Given that smoking may modify exposure, Eq. (3) was expanded to introduce the interaction of smoking (coded as 1 and 0 otherwise) with $PM_{2.5}$ exposure, as

$$logit(p_{ij}) = \alpha_i + \beta z_{ij} + \gamma(z_{ij} \cdot s_{ij}) + \varepsilon_{ij} \quad (4)$$

where s_{ij} is smoking dummy (1 = smoker, 0 otherwise), and γ is the coefficient of smoking-related exposure modifier.

RESULTS

PM Prediction Using Satellite-Based AOD

To compute $PM_{2.5}$ estimates for the entire study domain (i.e. Cleveland MSA from 2000 to 2009), we developed an empirical relationship of 2 km AOD with the *in situ* measurement of PM (at EPA sites Table 1), and then extrapolated the model for all AOD data points to predict $PM_{2.5}$ within the spatiotemporal domain under study.

Utilizing $PM_{2.5}$ and AOD centered (as described in the methodology section) on the selected sites, $PM_{2.5}$ was regressed on the instrumented AOD with the control for temporal structure and seasonality. The regression analysis was implemented in STATA using *ivxtregress* with the cluster option for site- and day-specific random effect that corrects for site- and day-specific autocorrelation in the error term.[34] The model was run separately for each site (where PM data are recorded on the ground) and for all sites together. The results of the analysis are presented in Table 2a. We compared the performance of global model with the site-specific model. As evident in Table 2b, in the site-specific model the predicted mean (less bias) and variance are closer to the observed mean and variance values, and root mean square error of the site-specific model is lower than that of the global model. The site-specific model will be useful to develop time-series estimates at/around the site. However, the prediction domain will be very limited in terms of geographic coverage. Therefore, it can result in partial spatial coverage of air pollution. We used the global model for predicting $PM_{2.5}$ for all points for which AOD was available. The spatiotemporal resolution of these data was the same as that of the original AOD data, retrieved at 2 km spatial resolution. The temporal resolution of AOD is once a day, that is, the overpass time that is fraction of a

second. Extrapolating the predictions (from the empirical model) resulted in a total of 2.34 million valid data points. On average, more than 120 000 $PM_{2.5}$ values were available for each year for each satellite within the geographic extent of the Cleveland MSA (82.5 °W—80.75 °W and 40.5 °N–41.92 °N). The number of data points doubled since 2003 within the same geographic extent after the launch of the Aqua satellite in May 2002, and the hourly extent of these data covered 0900–1500 hours local time. The average concentrations of $PM_{2.5}$ from Terra were significantly higher than that from Aqua (mean difference was ~1.42 μg/m^3, P 0.001). Explaining these differences requires more research. Nonetheless, one potential explanation underlying these differences can be attributed to changes in $PM_{2.5}$ concentration due to traffic during peak and off-peak hours. The local overpass time of Terra corresponds with the peak traffic hours, and Aqua overpass time corresponds with the off-peak hours. Thus, $PM_{2.5}$ estimates from Terra (peak hours) and Aqua (off-peak hours) combined can provide representative estimates of daily $PM_{2.5}$ concentration at a given location.

To evaluate the performance of LTSK, we created artificial gaps in the predicted data set. Given the size of data and computational constraints, we randomly chose 2.5% (33 480 data points of 1 347 407 data points between 2006 and 2009 in the Cleveland MSA) of the data points and then dropped them from the input data set supplied for computing LTSK (Table 3a). Using the remaining data points in the input data set, we computed $PM_{2.5}$ for (location and time of) the skipped data points with the aid of LTSK. We used 0.05° (~5 km) distance and 2 day time intervals to search for sample points needed for LTSK predictions. The mean and 95% confidence intervals of $PM_{2.5}$ (predicted using LTSK) and observed $PM_{2.5}$ are very close (Table 3b), and the correlation between them (when Kriging flag was Good) was ~0.9 (Table 3b and Figure 2). However, the correlation between them dropped to ~0.73 when fewer neighbors were found within the specified time interval. This means that the availability of sufficient number of data points (more than three data points) within the chosen spatiotemporal intervals (used for computing LTSK) is important to compute reliable exposure at a given location and time.

LTSK for Computing Exposure Estimates

For demonstration purposes, we overlaid a systematic grid of 2.5 km spatial resolution over Cleveland MSA and computed the daily $PM_{2.5}$ at the centroid of each 2.5 km cell from 2000 to 2009 (Table 3a). This resulted in a total of 2.13 million query points for which $PM_{2.5}$ were needed. To predict estimate at a given location and time, we need sample data points adjacent to this location and time. As mentioned earlier, a balance needs be made to define distance and time intervals so that there are sufficient data points around the query point (for which prediction is needed) to estimate covariance function. The LTSK, the program we have developed, outputs the prediction flags for each predicted value: good, less neighbors by space and time, less neighbor by time and missing values. The good prediction shows that there were sufficient data points for predicting estimates at a given location and time. In this example, we used 15 days and 0.1° to estimate the parameters of LTSK. All predicted values were evaluated for their robustness. Of these 2.13 million query points, LTSK predicted robust value (i.e. with "Good" flag) for 1.7 million (~79.8%) query points. Summary of predicted values is presented in Table 3b. Examples of predictions of $PM_{2.5}$ for three consecutive days in two different seasons are shown in Figure 3. As evident from this figure, there is greater spatiotemporal heterogeneity in $PM_{2.5}$ distribution. For example, across two consecutive days the value of $PM_{2.5}$ ranges between <10 and >40 μg/m^3. A major strength of LTSK is that it can capture and characterize spatio-temporal heterogeneity within narrower distance and time intervals, which otherwise is not possible using the global Kriging model and other traditional methods of interpolation.

AECOPD and time–space lagged exposure

AECOPD is a product of increase in inflammation because of viral, bacterial or environmental loading.[45] However, exacerbations vary in the degree of severity, with only severe exacerbations requiring an admission. Aaron et al.[46] reported the natural history of a COPD exacerbations, specifically describing the time course and pattern of onset. The authors followed 212 COPD subjects for 2.8 years. During this time interval they collected self-reported symptoms. Their analysis consisted of event time intervals. They concluded that AECOPD exhibits two distinct patterns: sudden onset and gradual onset. Although the sudden onset was associated with increased respiratory symptoms, it had a faster recovery time. Therefore, the lag time between exposure and hospital admission seems to suggest that the hospitalizations because of AECOPD were more gradual. Furthermore, exposures resulting in gradual onset of AECOPD could prolong the recovery time. Therefore, patients are more likely to seek medical attention at the hospital, ultimately requiring an admission. The mechanism of ambient air pollution induced respiratory exacerbation is unclear, but several reports suggest a direct proinflammatory effect of PM on macrophages.[47] Furthermore, PM interacts with lung epithelial cells, which can induce cell death and cell epithelial disruption.[48,49] All important mechanisms that impair innate immunity, promoting a proinflammatory environment, increase the risk of a host to acquire infections.

A total of 369 veterans were admitted for AECOPD during 2007; of these 39% were smokers and 10% had an admission for an AECOPD in the preceding year. Overall, the patient sample was generally older, male, with good representation of non-white race/ethnicity (Table 4). We performed two sets of analyses. In the first set, we modeled relative risks of exposed group (defined as $PM_{2.5}$ exposure >15.4 μg, which is 75% percentile of computed exposure values) with respect to unexposed group. In the second set, we modeled relative risks with respect to continuous $PM_{2.5}$ exposure. Results of the analyses are presented in Tables 5a–c.

We organize results across four distance intervals and five time lags. Exposure uncertainty tends to increase with the increase in spatial intervals.[8] However, time lag is associated with the incubation period (or harvesting period, that is, the time difference between exposure and physiological response). Restricting the exposure within very narrow spatiotemporal lags, such as 3 days and 0.025° (~2.5 km), exposure was computed for only 156 (~21%) observations (which included 71 cases and 85 controls). However, expanding the analysis to 0.05° (~5 km) and 6 days, we were able to compute exposure for 47% of the observations (348 = 181 cases and 167 controls) (Table 5a).

Our analysis suggests that if exposure computation was restricted within 0.05° (~5 km) and 6 days, exposed group has 54% greater chance of hospital admission because of AECOPD and the risk further increases within 9 and 12 day lags. Our results suggest that the incubation period of COPD exacerbation with regard to short-term exposure to ambient $PM_{2.5}$ is about 6–9 days. Similar findings emerge in the analysis of continuous $PM_{2.5}$ exposure without (Table 5b) and with (Table 5c) the smoking as PM effect modifier. For example, a unit increase in $PM_{2.5}$ exposure (computed using the data within 0.05° and 9 days) was associated with 2.5% increase in the risks of hospital admission because of AECOPD (Table 5b). Smoking is one of the major risk factors for an exacerbation, and a change in smoking behavior (especially when cases serve their controls) may confound the results. Therefore, we control for the effects of smoking as the $PM_{2.5}$ effect modifier. Although the coefficients of the effect modifier were insignificant for all distance and time lags, risks associated with $PM_{2.5}$ increased further within 15 days of exposure and 0.05° (~5 km) distance lag and the risks also became significant within 7.5 and 10 km distance lags and within 6, 9 and 12 days intervals (Table 5c). In addition, we also analyzed data for non-smokers separately (Tables 6a and b). Obviously, the sample size reduced significantly in

the analysis of non-smokers. In the aggregated data set (of exposed and reference groups), time–space lagged exposure did not show any significant risks of AECOPD (Table 6a). However, in the analysis of continuous $PM_{2.5}$ exposure, the results are consistent with results of entire group. For example, a unit increase within 9 days and 0.05° lags the risk of AECOPD increases by 2.6%; the value for the entire group was 2.5% (Table 6b).

Although our analysis shows that there is an increased risk of hospital admission because of AECOPD with respect to increase in $PM_{2.5}$ exposure, these findings should be interpreted with caution. We used a case crossover design to control for potential confounders. In this design, we assume that indoor exposure and exposure at other places (where these individuals might have spent significant time) remained the same. Likewise, we also assume that behaviors, choices and other risks factors of these individuals remained the same when cases served their own controls 30 days before AECOPD. Another limitation is that most of the subjects are male, because phenotypes of COPD may vary across gender. Therefore, these finding may largely be applicable to male veterans.

DISCUSSION

This paper presents a hybrid approach to develop daily estimates of exposure to ambient $PM_{2.5}$ and demonstrates the application of this approach to investigate the risks of AECOPD with respect to time–space lagged $PM_{2.5}$ exposure. The hybrid approach is an important development for air pollution epidemiological studies and air quality surveillance and management. This approach can be used to develop ambient air pollution estimates at a given location and time. It capitalizes upon the strengths of satellite remote sensing, *in situ* air pollution monitoring and optimal methods of interpolation technique to develop exposure at a given location and time. MODIS data have daily global coverage. Using these data, daily AOD can be retrieved at various spatial resolution, as fine as 1 km.[14,28,50] Thus, these data offer unprecedented spatial resolution and temporal coverage. *In situ* monitoring of air pollution at EPA sites provides opportunity to develop and validate air pollution prediction model with the aid of satellite-based AOD and meteorological conditions. LTSK is useful to quantify exposure (at a given location and time) with minimal prediction error.[19] It addresses the problems of misalignment, missing values and mismatch in the spatiotemporal resolutions of environment and health data sets.[19]

As evident from our analysis, the hybrid approach predicted estimates of daily $PM_{2.5}$ at 2.5 km spatial resolution for the Cleveland MSA from 2000 to 2009. Of the predicted values 80% were robust. This means within the study domain (i.e. Cleveland MSA and 2000 and 2009), we had 80% daily coverage of $PM_{2.5}$ data. The coefficient of correlation between observed and predicted (using LTSK) $PM_{2.5}$ was ~0.9. For the analysis of AECOPD, we were able to compute exposure at the place of residence for 55% of the data points when exposure estimation was restricted within 9 days and 0.05° (~5 km) distance. However, when exposure is computed using the *in situ* monitored EPA data, the sample size reduces dramatically. For example, in a recent research when exposure estimation was restricted within 3 miles distance from air pollution monitoring sites for the entire gestation (that included on an average 60 data points) the sample size dropped to ~10%.[8]

LTSK is a unique and novel development for exposure science, as it has the potential to quantify and characterize spatiotemporal heterogeneity, and identify the contribution of local emission sources. In addition, it addresses the issues of spatiotemporal misalignment, mismatch in the spatiotemporal scales of environment and health data sets. However, LTSK requires the search domain around a location and time where prediction is needed and large data size can pose computational problems. A balance needs to be made between the number of data points (i.e. sample size) and spatiotemporal intervals within which to search for

sample points for developing covariance function needed for LTSK. The narrower spatiotemporal intervals will result in a smaller sample size. However, wider spatiotemporal intervals can ensure fewer gaps in the prediction, but it leads to greater generalization and may fail to capture local spatiotemporal heterogeneity. Given the spatiotemporal heterogeneity and lifespan of aerosols (about 5 days), we recommend 3–5 daytime interval and ~5 km distance interval for LTSK estimation.

High-resolution PM data, computed using the hybrid approach, will enhance exposure estimation for population-based studies. Not only does this approach ensure greater precision in ambient air pollution exposure estimation but it can also be used to compute short-term air pollution exposure for population-based studies and can be extrapolated to other cities and larger geographic domains, especially where *in situ* monitoring of air pollution is absent. Air pollution estimates developed using this approach can also be useful to revisit air pollution epidemiological studies to improve our understanding of the burden of disease and disability associated with ambient air pollution. Because the hybrid approach ensures greater sample size (because of the increase in spatiotemporal coverage of air pollution) and reduces exposure misclassification. Thus, the results of these studies are likely to be more reliable than that of the traditional studies that rely on air pollution data from sparsely distributed fixed EPA sites. In addition, high-resolution air pollution, estimated using the hybrid approach, have greater implications for air quality surveillance and management and environmental justice research.

Although the hybrid approach seems promising to develop exposure, it is important to document inherent limitations of this approach. First, AOD retrieval is not possible every day because of cloud cover, cloud and contamination. Therefore, there are systematic gaps in AOD data. Second, the AOD–PM association can be influenced by a number of factors: emission sources and meteorological conditions and proximity natural sources of aerosol, such as sea salt, forest fires and windblown dust. Therefore, an AOD–PM association observed in one region cannot be directly extrapolated to other regions. It is also important to have *in situ* air pollution data to develop and validate AOD–PM empirical model. In the absence of such data, mobile vans can be used to record air pollution data instead of satellite over time. For example, Kumar and Foster[51] administered a mobile field campaign to collect air pollution data prospectively, restricting their monitoring during the satellite crossing time. Future efforts should be geared towards the assimilation of satellite-based AOD with the chemical transport models to further enhance the temporal coverage of air pollution prediction by hour.[50]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Davidson CI, Phalen R, Solomon P. Airborne particulate matter and human health: a review. Aerosol Sci Technol. 2005; 39:737–749.

2. Jia XF, Song XM, Shima M, Tamura K, Deng FR, Guo XB. Effects of fine particulate on heart rate variability in Beijing: a panel study of healthy elderly subjects. Int Arch Occup Environ Health. 2012; 85:97–107. [PubMed: 21573961]

3. Gualtieri M, Ovrevik J, Mollerup S, Asare N, Longhin E, Dahlman HJ, et al. Airborne urban particles (Milan winter-PM2. 5) cause mitotic arrest and cell death: effects on DNA, mitochondria, AhR binding and spindle organization. Mutat Res. 2011; 713:18–31. [PubMed: 21645525]

4. Murrell RN, Gibson JE. Brevetoxin 2 alters expression of apoptotic, DNA damage, and cytokine genes in Jurkat cells. Hum Exp Toxicol. 2011; 30:182–191. [PubMed: 20498034]

5. Nikasinovic L, Just J, Sahraoui F, Seta N, Grimfeld A, Momas I. Nasal inflammation and personal exposure to fine particles PM2. 5 in asthmatic children. J Allergy Clin Immunol. 2006; 117:1382–1388. [PubMed: 16751001]

6. O'Donnell R, Breen D, Wilson S, Djukanovic R. Inflammatory cells in the airways in COPD. Thorax. 2006; 61:448–454. [PubMed: 16648353]

7. Pope CA, Dockery DW, Kanner RE, Villegas GM, Schwartz J. Oxygen saturation, pulse rate, and particulate air pollution—a daily time-series panel study. Am J Resp Crit Care Med. 1999; 159:365–372. [PubMed: 9927345]

8. Kumar N. Uncertainty in the relationship between criteria pollutants and low birth weight in Chicago. Atmos Environ. 2012; 49:171–179.

9. Kumar N. What can affect AOD–PM2.5 association? Environ Health Perspect. 2010; 118:A109–A110.

10. Ott DK, Kumar N, Peters TM. Passive sampling to capture spatial variability in PM10–2. 5. Atmos Environ. 2008; 42:746–756.

11. Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J Expo Anal Environ Epidemiol. 2001; 11:231–252. [PubMed: 11477521]

12. Özkaynak H, Baxter L, Dionisio K, Burke J. Air pollution exposure prediction approaches used in air pollution epidemiology studies. J Expos Sci Environ Epidemiol. 2013 forthcoming.

13. Chu DA, Kaufman YJ, Ichoku C, Remer LA, Tanre D, Holben BN. Validation of MODIS aerosol optical depth retrieval over land. Geophys Res Lett. 2002:29.

14. Li C, Lau AK-H, Mao JT, Chu DA. Retrieval, validation and application of 1-km resolution aerosol optical depth from MODIS data over HongKong. Trans Geosc Remote Sens. 2005; 43:2650–2658.

15. Choi Y, Wang Y, Zeng T, Cunnold D, Yang ES, Martin R, et al. Springtime transitions of NO_2, CO, and O_3 over North America: model evaluation and analysis. J Geophys Res Atmos. 2008:113.

16. Huang M, Carmichael GR, Spak SN, Adhikary B, Kulkarni S, Cheng Y, et al. Multi-scale modeling study of the source contributions to near-surface ozone and sulfur oxides levels over California during the ARCTAS-CARB period. Atm Chem Phys. 2011; 11:3173–3194.

17. van Donkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C, et al. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. Environ Health Perspect. 2010; 118:847–855. [PubMed: 20519161]

18. Zhang Q, Streets DG, Carmichael GR, He KB, Huo H, Kannari A, et al. Asian emissions in 2006 for the NASA INTEX-B mission. Atmos Chem Phys. 2009; 9:5131–5153.

19. Liang D, Kumar N. Time–space Kriging to address the spatial misalignment, in the large datasets. Atmos Environ. 2013; 72:60–69.

20. Kumar N. What can affect AOD–$PM_{2.5}$ association? Environ Health Perspect. 2010; 118:A2–A3.

21. Schwartz J. Air pollution and hospital admissions for respiratory disease. Epidemiology. 1996; 7:20–28. [PubMed: 8664396]

22. Zanobetti A, Schwartz J. The effect of particulate air pollution on emergency admissions for myocardial infarction: a multicity case-crossover analysis. Environ Health Perspect. 2005; 113:978. [PubMed: 16079066]

23. Luís Ferreira Braga A, Zanobetti A, Schwartz J. The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities. J Occup Environ Med. 2001; 43:927. [PubMed: 11725331]

24. NASA. The Level 1 and Atmosphere Archive and Distribution System. National Aeronautics and Space Administration; Available from http://ladsweb.nascom.nasa.gov/ [3 July 2010]

25. NCDC. [Date last accessed 10 March 2011] National Climatic Data Center. 2011. Available from http://www.ncdc.noaa.gov/oa/ncdc.html

26. EPA. Envirofacts Data Warehouse. Environmental Protection Agency; Available from http://www.epa.gov/enviro/ [5 February 2008]

27. NASA. MODIS Atmosphere. Greenbelt, MD: NASA; Available from http://modisatmos.gsfc.nasa.gov/ [24 April 2013]

28. Kumar N, Chu AD, Foster AD, Peters T, Willis R. Satellite remote sensing for developing time and space resolved estimates of ambient particulate in Cleveland, OH. Aerosol Sci Technol. 2011; 45:1090–1108. [PubMed: 22238503]

29. Ramachandran S. Aerosol optical depth and fine mode fraction variations deduced from Moderate Resolution Imaging Spectroradiometer (MODIS) over four urban areas in India. J Geophys Res Atmos. 2007:112.

30. Gupta P, Christopher SA. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: multiple regression approach. J Geophys Res Atmos. 2009:114.

31. Tripathi SN, Srivastava ABK, Dey S, Satheesh SK, Krishnamoorthy K. The vertical profile of atmospheric heating rate of black carbon aerosols at Kanpur in northern India. Atmos Environ. 2007; 41:6909–6915.

32. Zhang JL, Reid JS. MODIS aerosol product analysis for data assimilation: assessment of over-ocean level 2 aerosol optical thickness retrievals. J Geophys Res Atmos. 2006:111.

33. NASA. The AERONET (AErosol RObotic NETwork). National Aeronautics and Space Administration; Available from http://aeronet.gsfc.nasa.gov/ [Date last accessed 10 March 2011]

34. StataCorp. STATA/SE Version 10.1. College Station, TX: StataCorp LP; 2010.

35. De Iaco S, Myers DE, Posa D. Space-time analysis using a general product–sum model. Stat Probab Lett. 2001; 52:21–28.

36. De Cesare L, Myers DE, Posa D. FORTRAN programs for space–time modeling. Comput Geosci UK. 2002; 28:205–212.

37. Zhang H, Wang Y. Kriging and cross-validation for massive spatial data. Environmetrics. 2010; 21:290–304.

38. Cressie N. Spatial prediction and ordinary Kriging. Math Geol. 1988; 20:405–421.

39. Kyriakidis PC, Journel AG. Geostatistical space-time models: a review. Math Geol. 1999; 31:651–684.

40. Vecchia AV. Estimation and model identification for continuous spatial processes. J R Stat Soc B Met. 1988; 50:297–312.

41. Rennen G. Subset selection from large datasets for Kriging modeling. Struct Multidiscip O. 2009; 38:545–569.

42. Cressie N, Shi T, Kang EL. Fixed rank filtering for spatio-temporal data. J Comput Graph Stat. 2010; 19:724–745.

43. Stevens DL, Olsen AR. Spatially balanced sampling of natural resources. J Am Stat Assoc. 2004; 99:262–278.

44. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991; 133:144–153. [PubMed: 1985444]

45. Decramer M, Janssens W, Miravitlles M. Chronic obstructive pulmonary disease. Lancet (Review). 2012; 379:1341–1351.

46. Aaron SD, Donaldson GC, Whitmore GA, Hurst JR, Ramsay T, Wedzicha JA. Time course and pattern of COPD exacerbation onset. Thorax. 2012; 67:238–243. [PubMed: 22008189]

47. Guastadisegni C, Kelly FJ, Cassee FR, Gerlofs-Nijland ME, Janssen NA, Pozzi R, et al. Determinants of the proinflammatory action of ambient particulate matter in immortalized murine macrophages. Environ Health Perspect. 2010; 118:1728–1734. [PubMed: 20663738]

48. Caraballo JC, Yshii C, Westphal W, Moninger T, Comellas AP. Ambient particulate matter affects occludin distribution and increases alveolar transepithelial electrical conductance. Respirology. 2011; 16:340–349. [PubMed: 21122029]

49. Urich D, Soberanes S, Burgess Z, Chiarella SE, Ghio AJ, Ridge KM, et al. Proapoptotic Noxa is required for particulate matter-induced cell death and lung inflammation. FASEB J. 2009; 23:2055–2064. [PubMed: 19237507]

50. Kumar N. A Hybrid approach for predicting $PM_{2.5}$ exposure. Environ Health Perspect. 2010:118.

51. Kumar N, Foster DA. Air quality interventions and spatial dynamics of air pollution in Delhi. Int J Environ Waste Manage. 2009; 4:85–111.
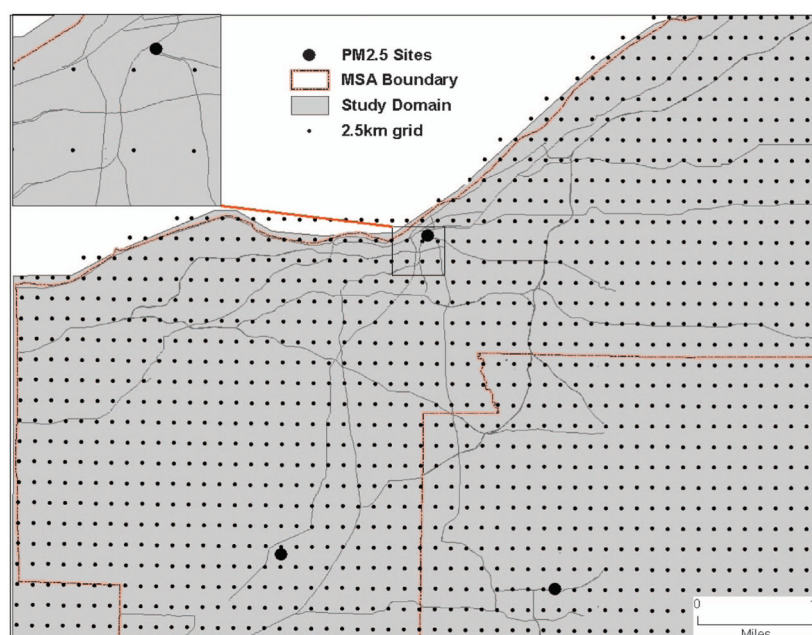
Figure 1.
Study area with the location of PM$_{2.5}$ (particulate matter 2.5) sites, with 2.5 km grid. The inset (right top) of the figure shows that there is just one PM$_{2.5}$ site in the downtown area and the spacing between two consecutive grid cells.

Figure 2.
Predicted (using local time–space Kriging (LTSK)) and observed $PM_{2.5}$ (particulate matter 2.5, $\mu g/m^3$) of 2.5% randomly chosen data points in Cleveland Metropolitan Statistical Area (MSA), 2006–2009 ($R^2 \sim 0.84$).
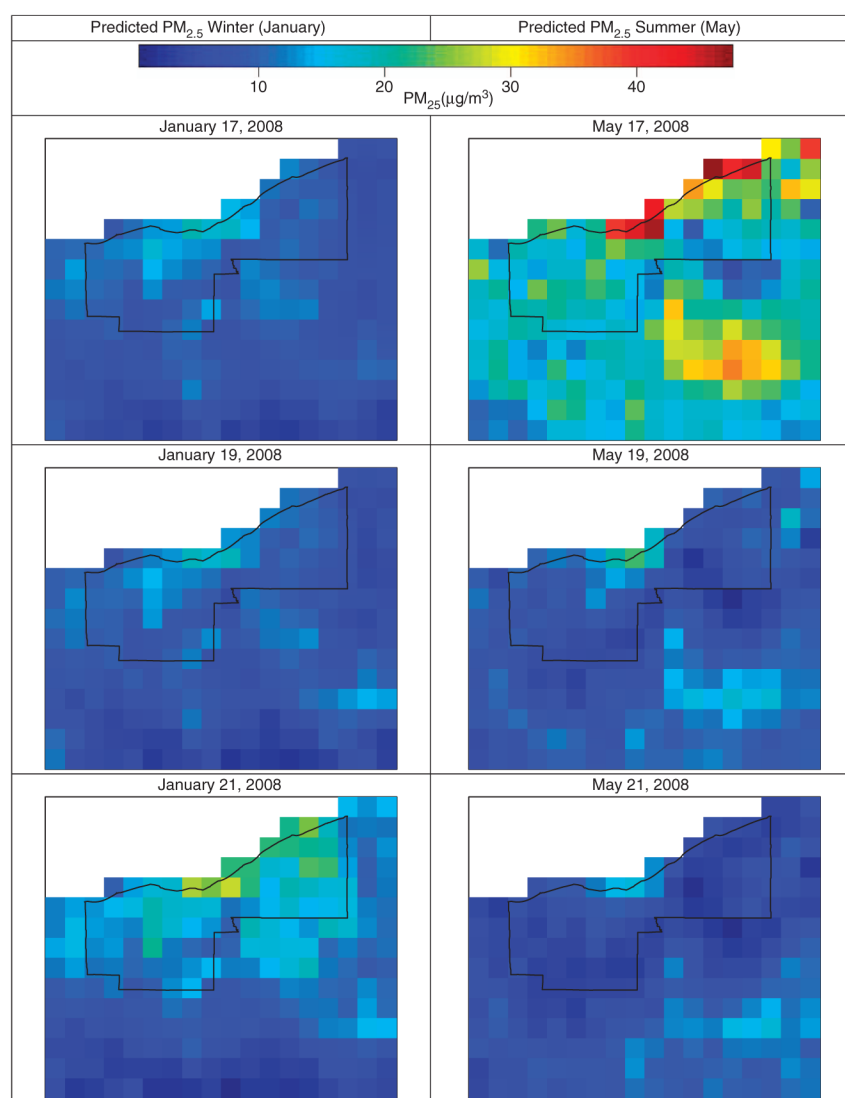
Figure 3.
PM$_{2.5}$ (particulate matter 2.5) predicted surface within Cleveland Metropolitan Statistical Area (MSA) across different seasons in 2008.

Table 1

PM$_{2.5}$ monitoring stations in Cleveland, OH.

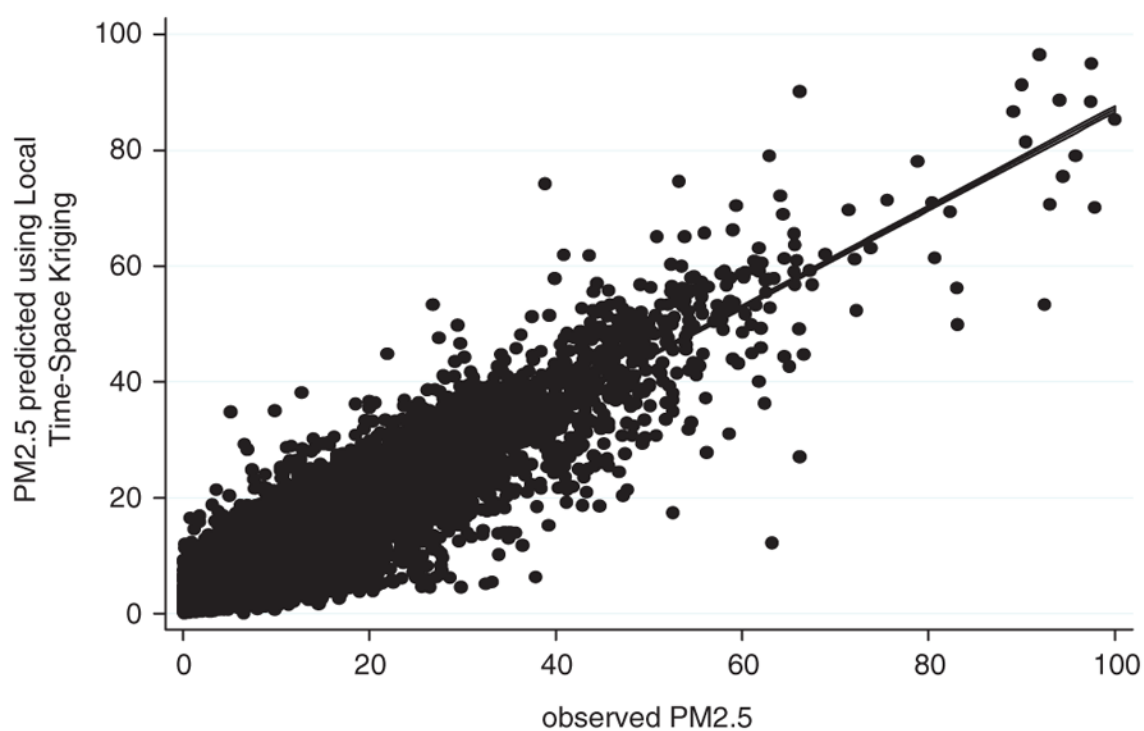| Site no.	Longitude (deg.)	Latitude (deg.)	Mean±95% confidence interval (μg/m³)	Operation beginning (year)
3	−81.9117	41.10278	16.1±0.11	2005
7	−80.7875	41.21417	13.1±0.11	2005
17	−81.4686	41.06333	14.1±0.10	2002
20	−81.3733	40.80056	10.5±0.07	2002
60	−81.6785	41.49395	16.0±0.13	2002
3002	−81.7005	41.58614	12.5±0.08	2004
Total	−81.5079	41.20068	13.3±0.04	NA

Abbreviations: EPA, Environmental Protection Agency; NA, not available; PM$_{2.5}$, particulate matter 2.5.

Source: EPA, 2010.

Table 2a

Site-specific instrumental regression estimates for PM$_{2.5}$.

Covariates	Site number					
	3	**7**	**17**	**20**	**60**	**Global**
Ln(AOD) = f (factors 1–3))	0.512*** (0.400–0.624)	0.898*** (0.686–1.110)	1.107*** (0.886–1.328)	1.364*** (1.090–1.638)	1.721*** (1.223–2.219)	0.982*** (0.865–1.099)
Sin(month no. since January 2000/6)	−0.222*** (−0.353—0.091)	−0.238*** (−0.406—0.070)	−0.178** (−0.334—0.022)	−0.397*** (−0.634—0.160)	0.04 (−0.242–0.322)	−0.269*** (−0.378—0.160)
Cos(month no. since January 2000/6)	0.0265 (−0.179–0.232)	0.176 (−0.094–0.446)	0.344*** (0.121–0.567)	0.445*** (0.131–0.759)	0.768*** (0.237–1.299)	0.232*** (0.065–0.399)
Season dummy (1 for June, July and Aug, 0 for the rest)	0.0237 (−0.188–0.235)	−0.108 (−0.402–0.186)	0.0443 (−0.250–0.338)	−0.095 (−0.454–0.264)	−0.0356 (−0.453–0.382)	−0.072 (−0.248–0.104)
Constant	3.864*** (3.595–4.133)	4.084*** (3.619–4.549)	4.385*** (3.975–4.795)	4.501*** (3.944–5.058)	4.813*** (4.127–5.499)	4.178*** (3.945–4.411)
Observations	286	236	395	301	212	1430

Abbreviations: AOD, aerosol optical depth; PM$_{2.5}$, particulate matter 2.5.

Factor 1 = a composite score of positively correlated surface temperature, dew point, relative humidity and negative visibility; Factor 2 = a composite score relative humidity inversely associated with surface temperature; and Factor 3 = a composite score of atmospheric pressure inversely associated with wind speed.

Source: Kumar et al.[33]

Table 2b

Validation and cross-validation of global- and site-specific models.

| Site number | Observed | Validation method | | | |
| | | Leave one out cross-validation | | 50% Sample and 50% control data set | |
		Global model	Site-specific model	Global model	Site-specific model
3	15.692±1.070 (297)	7.140±1.141 (296; 3.616)	16.098±1.074 (296; 1.842)	7.605±1.190 (148; 3.251)	17.151±1.092 (148; 1.777)
7	8.743±1.095 (250)	8.332±1.117 (249; 2.172)	8.993±1.108 (249; 2.102)	7.926±1.178 (116; 2.100)	8.383±1.181 (116; 2.132)
17	10.879±1.088 (405)	10.984±1.087 (406; 2.342)	10.772±1.099 (406; 2.493)	10.260±1.129 (226; 2.424)	9.500±1.144 (226; 2.585)
20	6.082±1.097 (545)	10.098±1.080 (545; 2.960)	6.593±1.113 (545; 3.102)	9.102±1.119 (283; 2.910)	5.319±1.166 (283; 3.178)
60	12.791±1.107 (218)	21.163±1.085 (222; 2.399)	12.971±1.147 (222; 2.598)	22.230±1.122 (114; 2.360)	14.614±1.218 (114; 2.702)
All sites	9.528±1.046 (1,715)	10.384±1.047 (1,718; 2.741)	9.858±1.052 (1,718; 2.535)	10.031±1.067 (887; 2.668)	9.060±1.079 (887; 2.601)

Abbreviations: PM$_{2.5}$, particulate matter 2.5; RMSE, root mean square error.

PM concentration ± 95% confidence interval ($\mu g/m^3$) The number of data points used are shown within parenthesis: and z_i is the observed value of PM$_{2.5}$.

$$\text{RMSE} = \sqrt{\sum_{i \in n} (\hat{z}_i - z_i)^2 / n}$$ where $\hat{z}_i$ is the estimated (or predicted value) PM$_{2.5}$

Note: Analyses were performed at log scale; values presented here are at the original scale.

Table 3a

Quality of PM$_{2.5}$ prediction for 2.13 million query points in Cleveland MSA, 2000–2009.

Prediction quality flag	Frequency	Percent
Good	1,701,262	79.78
Less neighbors by space time	27,426	1.29
Less neighbors by time	253,828	11.9
No neighbors found	142,289	6.67
Not a valid prediction	7655	0.36
Total	2,132,460	100

Abbreviations: MSA, Metropolitan Statistical Area; PM$_{2.5}$, particulate matter 2.5.

Table 3b

Validation of LTSK: predicted and observed $PM_{2.5}$ (mean±95% confidence interval ($\mu g/m^3$)).

Prediction flag	PM$_{2.5}$ predicted using LTSK	Observed PM$_{2.5}$ (derived using the empirical model[*])	Coefficient of correlation between Krigged and observed PM$_{2.5}$
Good	11.57±0.188 (12,780)	11.59±0.203 (12,780)	0.9194[**]
Less neighbor by space	NA	27.24±13.495 (13)	NA
Less neighbor by time	10.30±0.107 (20,674)	10.40±0.146 (20,674)	0.7375[**]
Non-computable	NA	2.51±1.446 (13)	NA
Total	10.79±0.098 (33,454)	10.86±0.119 (33,480)	NA

Abbreviations: LTSK, local time–space Kriging; NA, not available; PM$_{2.5}$, particulate matter 2.5.

[*]
Global model in Table 2a.

[**]
Significant at 1%.

Table 4

Descriptive statistics of veterans admitted for acute exacerbation of COPD in Cleveland, MSA, 2007.

Variable	N	%
Total	369	100
Gender		
Male	353	95.66
Female	16	4.34
Marital status		
Divorced	165	44.72
Other	1	0.27
Widow	42	11.38
Married	127	34.42
Never married	34	9.21
Income ($)		
5000	83	22.49
5001 to 10,000	63	17.07
10,001 to 15,000	81	21.95
15,001 to 25,000	67	18.16
25,001 or more	75	20.33
Urban residential status		
Yes	316	85.64
No	53	14.36
Race/ethnicity		
White	237	64.23
Black	112	30.35
Hispanic	4	1.08
Other	4	1.08
Unknown	12	3.25
Age (years)		
<45	<45	<45
45–55	45–55	45–55
55–65	−65	55–65
65–75	65–75	65–75
75–85	75–85	75–85
85 +	85 +	85 +
Admission for COPD in the past year		
Yes	39	10.57
No	330	89.43
Smoker		
Yes	144	39.02
No	225	60.98

Abbreviations: COPD, chronic obstructive pulmonary disease; LTSK, local time–space Kriging; MSA, Metropolitan Statistical Area.

Table 5a

Relative risk of hospital admission because of AECOPD with respect to exposed and unexposed groups in Cleveland MSA, 2007.

Lag (days)	0.025° (~2.5 km)	0.05°	0.075°	0.1°
<3	0.556 (156)	1.133 (254)	1.095 (303)	1.333 (335)
<6	1.000 (248)	1.542* (348)	1.235 (387)	1.405 (405)
<9	1.150 (309)	1.760** (391)	1.429 (418)	1.647** (427)
<12	1.192 (349)	1.769** (411)	1.424 (424)	1.471* (433)
<15	1.357 (375)	1.633** (418)	1.455* (428)	1.316 (433)

Abbreviations: AECOPD, acute exacerbation of COPD; COPD, chronic obstructive pulmonary disease; LTSK, local time–space Kriging; MSA, Metropolitan Statistical Area; PM$_{2.5}$, particulate matter 2.5.

P<|t|;

***
<0.01;

**
<0.05;

*
<0.10.

Note: PM$_{2.5}$ exposure was computed using LTSK, and exposed group was defined as PM$_{2.5}$ >15.4 μg/m^3.

Table 5b

Relative risk of hospital admission because of AECOPD with respect to time–space lagged $PM_{2.5}$ exposure (as continuous variable).

Lag (days)	0.025° (~2.5 km)	0.05°	0.075°	0.1°
<3	0.997 (156)	1.005 (254)	1.006 (303)	1.010 (335)
<6	1.000 (248)	1.012 (348)	1.009 (387)	1.012 (405)
<9	1.010 (309)	1.025[**] (391)	1.012 (418)	1.013 (427)
<12	1.007 (349)	1.024[**] (411)	1.012 (424)	1.012 (433)
<15	1.011 (375)	1.028[**] (418)	1.013 (428)	1.011 (433)

Abbreviations: AECOPD, acute exacerbation of COPD; COPD, chronic obstructive pulmonary disease; $PM_{2.5}$, particulate matter 2.5.

P<|t|;

[***] 0.01;

[**] <0.05;

[*] <0.10.

Table 5c

Relative risk of hospital admission because of AECOPD with respect to time–space lagged $PM_{2.5}$ exposure (as continuous variable) controlling for $PM_{2.5} \times$ smoking interaction (as the $PM_{2.5}$ exposure modifier due to smoking).

Lag (days)	0.025° (~2.5 km)	0.05°	0.075°	0.1°
<3	1.006 (±0.057,156)	1.003 (±0.029,254)	1.009 (±0.023,303)	1.013 (±0.02,335)
<6	1.00 (±0.033,248)	1.011 (±0.023,348)	1.014 (±0.019,387)	1.016[*] (±0.018,405)
<9	1.01 (±0.025,309)	1.025[**] (±0.023,391)	1.017[*] (±0.019,418)	1.017[*] (±0.018,427)
<12	1.007 (±0.022,349)	1.025[**] (±0.022,411)	1.017[*] (±0.02,424)	1.017[*] (±0.019,433)
<15	1.01 (±0.022,375)	1.029[**] (±0.023,418)	1.017[*] (±0.02,428)	1.015 (±0.019,433)

Abbreviations: AECOPD, acute exacerbation of COPD; COPD, chronic obstructive pulmonary disease; $PM_{2.5}$, particulate matter 2.5.

P<|t|;

[***] 0.01;

[**] <0.05;

[*] <0.10.

Table 6a

Non-smokers—relative risk of hospital admission because of AECOPD with respect to exposed and unexposed groups in Cleveland MSA, 2007.

Lag (days)	0.025° (~2.5 km)	0.05°	0.075°	0.1°
<3	0.375 (102)	0.909 (160)	0.733 (183)	1.000 (206)
<6	0.889 (156)	1.533 (215)	1.042 (240)	1.148 (251)
<9	1.000 (194)	1.471 (243)	1.154 (260)	1.360 (266)
<12	0.944 (222)	1.556 (256)	1.160 (264)	1.240 (270)
<15	1.333 (237)	1.381 (262)	1.240 (268)	1.103 (270)

Abbreviations: AECOPD, acute exacerbation of COPD; COPD, chronic obstructive pulmonary disease; LTSK, local time–space Kriging; MSA, Metropolitan Statistical Area; PM$_{2.5}$, particulate matter 2.5.

$P<|t|$;

[***] <0.01;

[**] <0.05;

[*] <0.10. PM$_{2.5}$ exposure was computed using LTSK and exposed group was defined as PM$_{2.5}$ >15.4 μg/m^3.

Table 6b

Non-smokers—relative risk of hospital admission because of AECOPD with respect to time–space lagged PM$_{2.5}$ exposure (as continuous variable).

Lag (days)	0.025° (~2.5 km)	0.05°	0.075°	0.1°
<3	0.992 (102)	1.009 (160)	1.002 (183)	1.008 (206)
<6	1.005 (156)	1.015 (215)	1.005 (240)	1.007 (251)
<9	1.013 (194)	1.026** (243)	1.008 (260)	1.010 (266)
<12	1.006 (222)	1.022* (256)	1.008 (264)	1.008 (270)
<15	1.012 (237)	1.026** (262)	1.010 (268)	1.008 (270)

Abbreviations: AECOPD, acute exacerbation of COPD; COPD, chronic obstructive pulmonary disease; PM$_{2.5}$, particulate matter 2.5.

$P<|t|$;

***
0.01;

**
<0.05;

*
<0.10.